Structural flexibility in the *Burkholderia mallei* genome

William C. Nierman*†‡, David DeShazer§, H. Stanley Kim*, Herve Tettelin*, Karen E. Nelson*, Tamara Feldblyum*, Ricky L. Ulrich§, Catherine M. Ronning*, Lauren M. Brinkac*, Sean C. Daugherty*, Tanja D. Davidsen*, Robert T. Deboy*, George Dimitrov*, Robert J. Dodson*, A. Scott Durkin*, Michelle L. Gwinn*, Daniel H. Haft*, Hoda Khouri*, James F. Kolonay*, Ramana Madupu*, Yasmin Mohammoud*, William C. Nelson*, Diana Radune*, Claudia M. Romero*, Saul Sarria*, Jeremy Selengut*, Christine Shamblin*, Steven A. Sullivan*, Owen White*, Yan Yu*, Nikhat Zafar*, Liwei Zhou*, and Claire M. Fraser*¶‖

*Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850; Departments of †Biochemistry and Molecular Biology, ¶Pharmacology, and ‖Microbiology and Tropical Medicine, George Washington University School of Medicine, 2300 Eye Street NW, Washington, DC 20037; and §U.S. Army Medical Research Institute for Infectious Diseases, 1425 Porter Street, Fort Detrick, MD 21702-5011

Edited by E. Peter Greenberg, University of Iowa, Iowa City, IA, and approved July 26, 2004 (received for review May 10, 2004)

The complete genome sequence of *Burkholderia mallei* ATCC 23344 provides insight into this highly infectious bacterium's pathogenicity and evolutionary history. *B. mallei*, the etiologic agent of glanders, has come under renewed scientific investigation as a result of recent concerns about its past and potential future use as a biological weapon. Genome analysis identified a number of putative virulence factors whose function was supported by comparative genome hybridization and expression profiling of the bacterium in hamster liver *in vivo*. The genome contains numerous insertion sequence elements that have mediated extensive deletions and rearrangements of the genome relative to *Burkholderia pseudomallei*. The genome also contains a vast number (>12,000) of simple sequence repeats. Variation in simple sequence repeats in key genes can provide a mechanism for generating antigenic variation that may account for the mammalian host's inability to mount a durable adaptive immune response to a *B. mallei* infection.

The etiologic agent of the disease known as glanders, *Burkholderia mallei*, has come under renewed scientific investigation as a result of recent concerns about biological weapons of mass destruction. As a Centers for Disease Control and Prevention category B agent, *B. mallei* is regarded as a potential biological weapon because it is highly infectious as an aerosol and results in a disease that is painful, incapacitating, difficult to diagnose, and often fatal (1). *B. mallei* is a highly evolved obligate parasite of horses, mules, and donkeys with no other known natural reservoir (1–3). Glanders is one of the oldest diseases known and was first described by Aristotle (384–322 before Christ) (4). *B. mallei* was used as a biological weapon during the American Civil War (4), World War I (5), and World War II and was reported to be weaponized by the Soviet Union for use as a biological weapon in Afghanistan. Here, we report the complete genome sequence of *B. mallei* ATCC 23344, a highly pathogenic clinical isolate and type strain of the species (6, 7) (see *Supporting Text*, which is published as supporting information on the PNAS web site). Bioinformatic and laboratory analysis of the genome provides further insight into the pathogenesis and biology of *B. mallei* and its relationship to pathogenic *Burkholderia pseudomallei* and nonpathogenic *Burkholderia thailandensis*.

Materials and Methods

Sequencing. The genome of *B. mallei* ATCC 23344 was sequenced and assembled by using the random shotgun method (8).

Coding Sequence (CDS) Prediction and Gene Identification. ORFs likely to encode proteins (CDSs) were identified by using GLIMMER (9). Identified CDSs were annotated by manual curation of the outputs of a variety of similarity searches. Searches of the predicted coding regions were performed with BLASTP, as described (10). The protein–protein matches were aligned with blast_extend_repraze, a modified Smith-Waterman (11) algorithm that maximally extends regions of similarity across frameshifts. Gene identification is facilitated by searching against a database of nonredundant bacterial proteins (nraa) developed at The Institute for Genomic Research (TIGR) and curated from the public archives GenBank, Genpept, Protein Information Resource, and SwissProt. Searches matching entries in nraa have the corresponding role, gene common name, percent identity and similarity of match, pairwise sequence alignment, and taxonomy associated with the match assigned to the predicted coding region and stored in the database. CDSs were also analyzed with two sets of hidden Markov models constructed for a number of conserved protein families from PFAM (12) and TIGRFAM (13). Regions of the genome without CDSs and CDSs without a database match were reevaluated by using BLASTX as the initial search, and CDSs were extrapolated from regions of alignment. Finally, each putatively identified gene was assigned to one of 113 role categories adapted from Riley (14).

Trinucleotide Signatures Analysis. The data for trinucleotide signature analysis were computed by the method described by Karlin (15) with a window size of 100,000 bp and a granularity of 100 bp. For χ^2 analysis, the distribution of all 64 trinucleotides (3-mer) was computed for the complete genome in all six reading frames, followed by the 3-mer distribution in 2,000-bp windows. Windows overlapped by 1,000 bp. For each window, the χ^2 statistic on the difference between its 3-mer content and that of the whole genome was computed.

Microarray Methods. Whole-genome DNA microarrays for *B. mallei* were fabricated as follows. DNA amplicons representing the genome were amplified, purified, and printed on Corning UltraGAPS aminosilane-coated microscope slides by using a robotic spotter built by Intelligent Automatic Systems (Cambridge, MA) and cross-linked by UV illumination.

Genomic DNA was prepared from 10 *B. mallei* isolates (see Fig. 2) by using a DNeasy kit (Qiagen, Valencia, CA). Purified genomic DNA was labeled and hybridized as described (16). Total RNA from the livers of three Syrian golden hamsters

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CDS, coding sequence; TIGR, The Institute for Genomic Research; cfu, colony-forming units; IS, insertion sequence; SSR, simple sequence repeat.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. CP000010 and CP000011).

‡To whom correspondence should be addressed. E-mail: wnierman@tigr.org.

Table 1. *B. mallei* genome features

ORF summary	
Total ORFs	5,535
Assigned function	3,173
Conserved hypothetical	718
Unknown function	564
Hypothetical proteins	978
Chromosome 1	
Size, bp	3,510,148
No. of ORFs	3,344
Percent coding	87.1
Average gene length, nt	914
GC content, %	68.2
rRNA operons	1 complete, 1 16S only
tRNAs	47
Chromosome 2	
Size, bp	2,325,379
No. of ORFs	2,091
Percent coding	85.7
Average gene length, nt	953
GC content, %	69.0
rRNA operons	2 complete
tRNAs	9

infected for 2 days was isolated by using TRIzol (Life Technologies, Rockville, MD) following the manufacturer's protocol. Labeling reactions with RNA and hybridization were conducted as described in the TIGR standard operating procedures found at http://atarrays.tigr.org. All of the hybridizations between two samples were replicated in dye-swap sets to obtain robust data.

Hybridized slides were scanned by using an Axon Instruments (Union City, CA) GenePix 4000B microarray scanner, and the independent images from each channel were analyzed by using TIGR SPOTFINDER (www.tigr.org/software) to obtain relative transcript levels. Data from TIGR SPOTFINDER were stored in AGED, a relational database designed to effectively capture and store microarray data. Data were normalized by using the MIDAS software tool (www.tigr.org/software). All calculated gene expression ratios were $\log_2$-transformed and averaged over dye-swapped arrays. The resulting data were visualized and further explored by using the TIGR data mining package (www.tigr.org/software).

Animal Studies. Syrian hamsters (five per group) were infected i.p. with 10^1, 10^2, and 10^3 colony-forming units (cfu) of each *B. mallei* strain. Deaths were monitored for 7 days, and the LD_{50} values were determined. The LD_{50} values for virulent and avirulent *B. mallei* strains were $<10^1$ cfu and $>10^3$ cfu, respectively. The LD_{50} for a *B. mallei* ATCC 23344 *virG* mutant was $>2.6 \times 10^3$ cfu (see *Supporting Text* for construction of virG mutant). Research was conducted in compliance with the Animal Welfare Act and other federal statutes and regulations relating to animals and experiments involving animals and adheres to principles stated in the *Guide for the Care and Use of Laboratory Animals*, National Research Council, 1996. The facility where this research was conducted is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International.

Results and Discussion

Genome Features. The genome of *B. mallei* consists of two circular chromosomes. Chromosome 1 contains 3,510,148 bp, and chromosome 2 contains 2,325,379 bp. A total of 5,535 predicted protein-encoding ORFs were identified in the genome (Table 1). GC skew analysis does not unambiguously reveal the origin of replication of either of the chromosomes (Fig. 1), unlike the case of the *B. pseudomallei* chromosomes. Nucleotide position 1 of chromosome 1 is assigned to the region flanked on one side by the *dnaA* (BMA0001) and *dnaN* (BMA0002) genes and on the other by *rnpA* (BMA3399). Clustered near the assigned nucleotide 1 position on chromosome 2 are the genes coding for a *parA* homolog (BMAA2114), a putative partitioning protein, *parB* homolog (BMAA2115), and a putative replication protein (BMAA2117), suggesting a plasmid-like replicon structure. Like the smaller chromosome of *B. pseudomallei* (17), chromosome 2 contains essential metabolic genes, making this replicon indispensable.

Insertion Sequences (ISs). The *B. mallei* genome harbors numerous IS elements that most likely have mediated extensive genomewide insertion, deletion, and inversion mutations relative to the *B. pseudomallei* K96243 (Fig. 1 and Table 3, which is published as supporting information on the PNAS web site). IS elements account for ≈3.1% of the genome in 171 complete and partial IS elements that are representatives of the IS*3*, IS*5*, IS*110*, IS*256*, and ISL*3* families. In comparison, the *B. pseudomallei* K96243 genome contains only 42 and the *B. thailandensis* E264 genome contains only 46 complete or partial copies of IS elements (data not shown). Five lines of evidence suggest that these IS elements have been instrumental in genome alteration of *B. mallei*. (*i*) Most of the synteny break points between the *B. pseudomallei* and *B. mallei* genomes are bounded by IS elements in *B. mallei*. Numerous sites of GC skew discontinuity are observed at these breakpoint boundaries in *B. mallei* (Fig. 1). (*ii*) Analysis of the genome sequence revealed disruption of 23 genes and 10 IS elements by the IS elements (Table 4, which is published as supporting information on the PNAS web site). (*iii*) Two syntenic portions of chromosome 1 in *B. pseudomallei*, one of 41 kb in size and another of 5 kb, found on chromosome 2 in *B. mallei* (Fig. 3*D*, which is published as supporting information on the PNAS web site), are flanked by IS elements. (*iv*) The *B. mallei* genome contains a *sacB* gene that confers sucrose sensitivity. When plated on a medium containing sucrose, the *sacB* gene is deleted in the surviving colonies as revealed by comparative genome hybridizations. The deletions in 9 of the 10 surviving isolates are variably bounded by four flanking IS elements (Fig. 4, which is published as supporting information on the PNAS web site). All of these isolates retain full virulence. (*v*) The integration of IS*407*A results in a 4-bp chromosome insertion site duplication flanking the inserted IS (18). However, most of the IS*407*A copies distributed throughout the *B. mallei* genome are not flanked by 4-bp duplications. This finding suggests that most of the IS elements went through homologous recombination events between copies of IS*407*A, and during the shuffling processes the IS elements lost their 4-bp duplication signatures for *de novo* transpositions. IS element-mediated recombination and *de novo* insertion have been and continue to be active in genome structure alteration.

Environmental Survival of *B. mallei*. The *B. mallei* genome is smaller (5.8 Mb) than that of *B. pseudomallei* (7.2 Mb) or *B. thailandensis* (6.7 Mb). Whereas these latter organisms are environmental soil inhabitants, previous studies have concluded that *B. mallei*, an obligate mammalian parasite, does not survive well in the environment (19). By comparison to the *B. pseudomallei* genome, we found that 627 genes on chromosome 1 and 819 on chromosome 2 of *B. pseudomallei* are either not present or variant in *B. mallei* (Table 5, which is published as supporting information on the PNAS web site), whereas the others closely match between the two genomes (i.e., >90% identity; average, 99.1%) over >90% of the length (average, 99.9%). Most of these *B. pseudomallei*-specific genes are

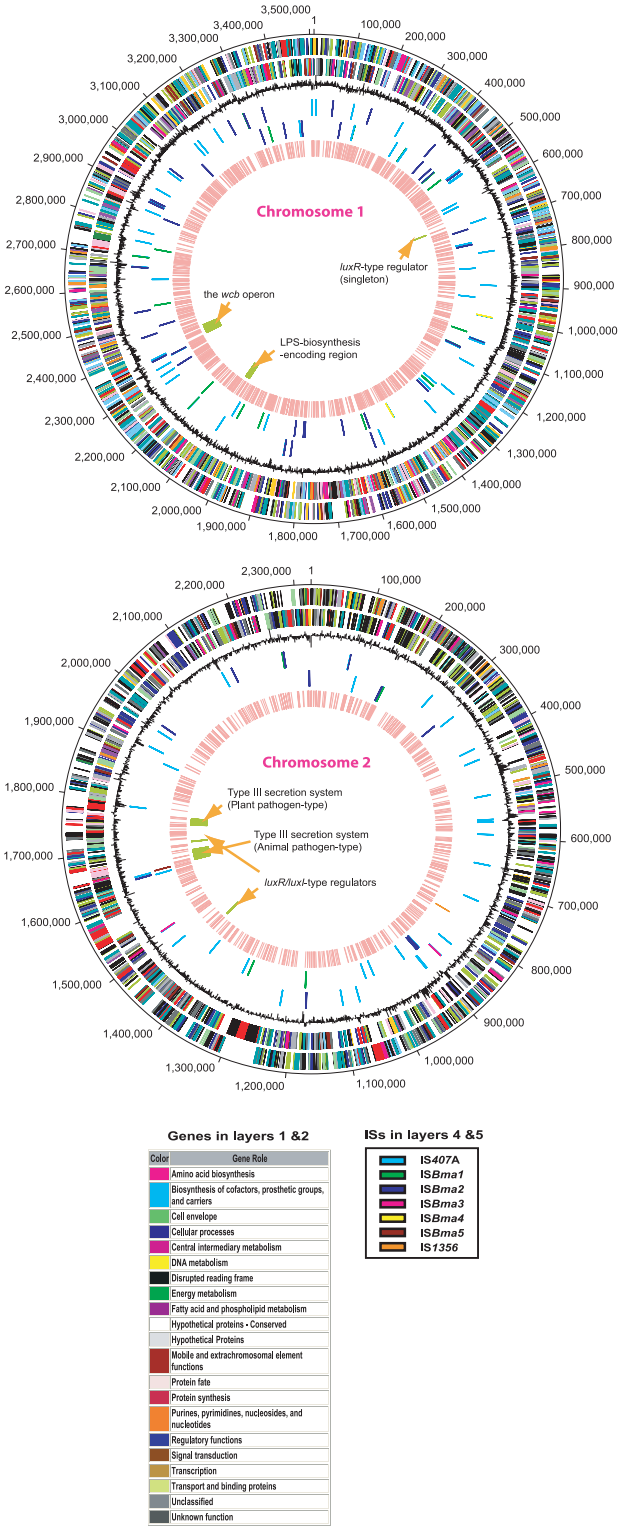

Fig. 1. Circular diagrams of chromosome 1 and chromosome 2 in *B. mallei*. Locations of selected genome features are denoted on seven layers of circles. From the outermost layer, 1) CDS in color codes for predicted functional role categories, 2) CDS in the other orientation, 3) GC skew, 4) IS elements flanking syntenic breaks compared to *B. pseudomallei* chromosome, 5) IS elements not associated with syntenic breaks, 6) SSRs that contain at least one A or T in the repeat unit, and 7) putative virulence genes. The locations of *wcb* capsule gene cluster, lipopolysaccharide biosynthesis genes, and a singleton *luxR*-type regulator in chromosome 1, plant and animal pathogen-type type III secretion system loci, and two pairs of *luxR*/*luxI*-type regulators in chromosome 2 are shown.

located in contiguous clusters of up to 150 kb in length (11), consistent with the IS-mediated deletions as the mechanism of gene loss in *B. mallei*. In addition, 37 *B. mallei* genes contain frameshifts relative to the *B. pseudomallei* orthologs (Table 6, which is published as supporting information on the PNAS web site), and another 10 contain transposon insertions (Table 4). These data coupled with experimental substrate utilization patterns as assayed by Biolog Phenotype Microarray technology (20) allowed the prediction of pathways specific to the metabolic capabilities of the *B. pseudomallei* relative to *B. mallei* (Fig. 5, which is published as supporting information on the PNAS web site). These observations suggest that metabolic capabilities essential for environmental survival may have been lost in the genome reduction process coincident with adaption of *B. mallei* to the mammalian host environment.

Simple Sequence Repeats (SSRs). The genome of *B. mallei* contains >12,000 SSRs within the CDS of 3,752 genes (68% of the genome), and in putative promoter regions of 179 genes. Of these 12,000 SSRs, 86.8% are located within coding regions and contain repeat units in multiples other than three, potentially leading to frameshift mutations upon replication error-induced changes in the number of repeat units in these genes (21, 22). Compared to other bacteria with similar GC content that have the same selection-independent statistical likelihood to contain G/C repeats, *B. mallei* has the highest density of SSRs by a nearly 2-fold margin (Table 2). The *B. pseudomallei* and *B. thailandensis* genomes also show a similar high density of SSR contents in their genomes (Table 2). Two analyses were performed to provide evidence that these SSRs may function as loci of altered protein expression or structure variation. (*i*) Annotation of the *B. mallei* genome identified 76 genes containing frameshifts relative to *B. pseudomallei* orthologs. By comparing these orthologs, 37 were found to contain frameshifting sequence differences in *B. mallei* caused by SSR repeat number differences between the orthologs (Table 6). (*ii*) The SSR loci from four genes coding for surface or putative virulence proteins were examined for variations in 39 strains of *B. mallei* and *B. pseudomallei*. Considerable variation in these loci was revealed not only between *B. mallei* and *B. pseudomallei* strains, but also within species (Fig. 6, which is published as supporting information on the PNAS web site). These instances of variation in SSR number are consistent with the SSR instability leading to phase variation in critical genes during the infection cycle.

Comparative Analysis. Comparative genome analysis has provided an opportunity to observe sequence differences that have arisen upon removal of selective pressure in the *B. mallei* chemotaxis and motility and the general secretory pathway genes. In contrast to *B. pseudomallei* and *B. thailandensis*, *B. mallei* is a nonmotile organism that lacks flagella (23, 24) yet its genome contains numerous flagellar and chemotaxis genes, most of which appear to be intact. One methyl-accepting chemotaxis gene has a frameshift mutation. In addition, a 65-kb insert flanked by IS elements disrupts the *fliP* gene, an essential gene for flagellum biogenesis (21), and a frameshift mutation in the flagellum motor *motB* gene eliminates its functionality. The mutations in these few pseudogenes likely account for *B. mallei* being nonmotile and nonflagellated, whereas the remainder of the chemotaxis and motility genes have not yet decayed.

In a similar fashion, *B. pseudomallei* has been characterized to be a strong secretor of proteases, lipases, and phospholipase C via the general secretory pathway (22), whereas *B. mallei* poorly secretes these proteins, if at all. The *B. mallei* genome analysis reveals that the genome contains the entire general secretory pathway cluster found in *B. pseudomallei*. Comparison of the 12

Table 2. SSR content of selected bacteria with %GC matching that of *B. mallei*

Bacteria	GC content, %	Length, bp	SSRs, all	
			No.	Density bp/SSR
Deinococcus radiodurans R1	66.9	2,648,638 (Chr 1)	2,665	993.86
Mycobacterium tuberculosis CDC1551	65.5	4,403,837	2,012	2,188.79
Bordetella pertussis Tohama I	67.6	4,086,189	4,873	838.54
Halobacterium sp. NRC-1	67.9	2,014,239	1,778	1,132.87
B. pseudomallei K96243	67.6	4,210,407 (Chr 1)	8,560	491.87
B. thailandensis E264	67.1	3,860,986 (Chr 1)	7,225	534.39
B. mallei ATCC 23344	68	3,510,148 (Chr 1)	7,084	495.50

SSRs examined in this comparison include homopolymers of at least six repeat units, di- and trinucleotide repeats of at least four units, and larger repeats up to heptamers of at least three units. Chr, chromosome.

B. mallei genes in this cluster to those of *B. pseudomallei* revealed that, whereas most of the genes are identical or nearly identical, the *B. mallei gspJ* gene contains a frameshift in the 3′ region of the gene, and the *gspL* gene contains a 54-base insertion for 18 aa. One or both differences are likely to account for the secretion phenotype differences between these organisms. The genes encoding chemotaxis and motility and the general secretory pathway in *B. mallei* are largely intact with only few mutations in few genes destroying the function of these pathways and accounting for the motility and secretory phenotype difference between *B. pseudomallei* and *B. mallei*.

Most *B. mallei* strains exhibit resistance to a number of antibiotics. The *B. mallei* genome contains at least 33 genes that may contribute to these drug resistances (Table 7, which is published as supporting information on the PNAS web site). *B. pseudomallei* is resistant to macrolide and aminoglycoside antibiotics via a multidrug efflux pump (25), whereas *B. mallei* is sensitive to these. Genes for this pump are located in a 50-kb region of chromosome 1 in *B. pseudomallei* that is absent in the *B. mallei* chromosome, thus accounting for this sensitivity.

Secondary Metabolite Genes. Soil bacteria, fungi, and some bacterial endosymbionts produce secondary metabolite compounds with antimicrobial and/or antimetabolite activities. In several instances, virulence depends on the production of a secondary metabolite with toxin activity (26). The *B. mallei* genome contains numerous (a total of 33) genes for nonribosomal peptide synthases and polyketide synthases that are located primarily in clusters on the smaller chromosome. These genes are a subset of those present in *B. pseudomallei* (Table 8, which is published as supporting information on the PNAS web site). These genes could code for biosynthetic enzymes for toxin production or they could be relics of its ancestral life as an environmental organism.

Virulence. *B. mallei* has not been extensively studied, and little is known about the virulence mechanisms of this bacterium. The only virulence factors definitely shown to be essential for pathogenicity are an extracellular capsule (18) and a *Salmonella typhimurium*-like type III secretion system (27) (see Fig. 1 for their locations in the chromosomes). The *B. mallei* genome encodes an extensive list of genes involved in regulating the levels of specific mRNAs. Among these is a DNA-binding response regulator (BMAA0762) with homology to the *virG* gene from *Agrobacterium tumifaciens* (28). It is located on chromosome 2 adjacent to its presumed cognate sensor histidine kinase. Insertional mutagenesis of this gene in *B. mallei* ATCC 23344 results in an increase in LD_{50} when tested for virulence in Syrian golden hamsters from $<10^1$ cfu to $>2.6 \times 10^3$ cfu, suggesting that this response regulator positively controls genes required for virulence. It is likely that other regulatory proteins will be shown to influence the *B. mallei* virulence phenotype.

As an initial effort to use the *B. mallei* genome sequence to identify the genes required by *B. mallei* to initiate and maintain pathogenesis, we conducted comparative genome hybridization analyses with 10 *B. mallei* strains, three of which have tested to be avirulent in Syrian golden hamsters (Fig. 2). Whereas most of the genes are shared across all of the isolates, we found 162 genes significantly diverged or absent in the avirulent isolates. These genes are present in both chromosomes, but there was a significant bias toward chromosome 2. Half of these showed little or no variation in the virulent isolates (Fig. 2, groups 1–3) and include 12 of the genes absent in the avirulent strains that are putative virulence genes based on their annotated function. Three additional genes missing in these isolates were shown to be expressed at a higher level in hamster liver than in LB medium (see below).

In a further effort to identify a set of genes that is likely to be essential for survival and proliferation of *B. mallei* in a mammalian host, expression profiling of the bacterial genome was performed by using total RNA isolated from the livers of three infected Syrian golden hamsters. We compared this gene expression in the host to that of an exponential growth phase *in vitro* culture in LB medium at 37°C by using a whole genome ORF *B. mallei* microarray. A set of 95 genes were found to be expressed at a higher level in the bacteria residing in the hamster livers than in the LB growth cultures, whereas 153 showed the opposite patterns based on the 95% confidence level (Table 9, which is published as supporting information on the PNAS web site). The less expressed genes in the hamster include those coding for cell cycle proteins and ribosomal proteins, reflecting the slower growth rate of the bacteria in the hamsters, and several iron storage proteins. The more expressed genes include those coding for secreted hydrolyses, a toxin biosynthesis protein, and phosphorus and iron acquisition proteins, reflecting the *in vivo* gene expression level adjustments to limitations for iron and phosphorous in the host, and metabolic adjustments necessary for the hamster environment. Higher expression levels of many known and putative virulence genes, such as type III secretion system components, adhesins, stress response genes, capsule, etc., were not observed in the hamster relative to LB culture, suggesting that these genes are not differentially expressed in culture and the hamster host or are regulated to be expressed at only specific times or under specific circumstances during infection.

Conclusion

The *B. mallei* genome sequence provides a starting point for studying the evolutionary history and the pathobiology of this organism. In evolving from a metabolically versatile soil organism to a highly specialized obligate mammalian pathogen, structural flexibility appears to be a major adaptive feature of its genome. This level of flexibility mediated through recombina-

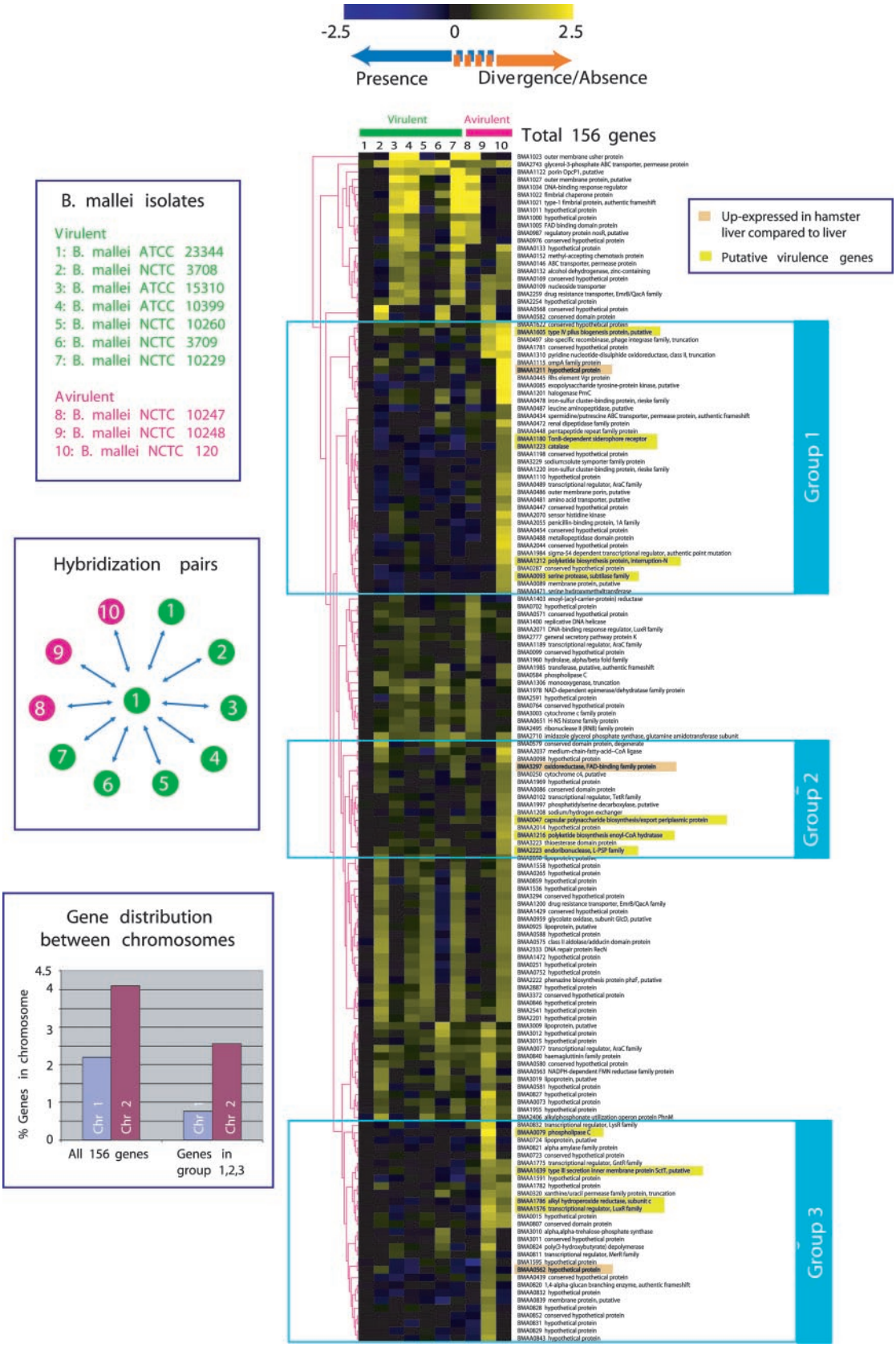

Fig. 2. Putative virulence genes revealed by comparative genome hybridization analysis. A total 156 genes diverged or absent in at least one of the three avirulent isolates were compared across the virulent isolates by using hierarchical clustering. Eighty genes that show little or no variation in virulent isolates were found in three different locations in the cluster, denoted as groups 1–3. Putative virulence genes and the genes shown to be highly expressed in hamster spleen (see Table 9) are highlighted. *B. mallei* isolates used for this study, microarray hybridization pairs, and a graph of gene distribution by chromosomes are also shown.

tion events among the numerous insertion elements and by means of simple sequence repeat-induced replication errors may be the mechanism used by this organism to effectively adapt to multiple distinct mammalian hosts (29) and to implement its extraordinary ability to evade the adaptive immune responses in the host.

We thank Julian Parkhill and colleagues at the Wellcome Trust Sanger Institute (Cambridge, U.K.) for sharing data before publication. This work was supported by National Institutes of Health/National Institute of Allergy and Infectious Disease Grants U01AI49037A, R01AI50565, and R01AI56006 (to W.C.N.).

1. Neubauer, H., Meyer, H. & Finke, E. J. (1997) *Rev. Int. Services Sante Forces Armees* **70,** 258–265.
2. McGilvray, C. D. (1944) *Can. J. Public Health* **35,** 268–275.
3. Alibasoglu, M., Yesildere, T., Calislar, T., Inal, T. & Calsikan, U. (1986) *Berl. Munch. Tierarztl. Wschr.* **99,** 57–63.
4. Sharrer, G. T. (1995) *Agric. Hist.* **69,** 79–97.
5. Christopher, G. W., Cieslak, T. J., Pavlin, J. A. & Eitzen, E. M., Jr. (1997) *J. Am. Med. Assoc.* **278,** 412–417.
6. Lopez, J., Copps, J., Wilhelmsen, C., Moore, R., Kubay, J., St-Jacques, M., Halayko, S., Kranendonk, C., Toback, S., DeShazer, D., *et al*. (2003) *Microbes Infect.* **5,** 1125–1131.
7. Srinivasan, A., Kraus, C. N., DeShazer, D., Becker, P. M., Dick, J. D., Spacek, L., Bartlett, J. G., Byrne, W. R. & Thomas, D. L. (2001) *N. Engl. J. Med.* **345,** 256–258.
8. Nierman, W. C., Feldblyum, T. V., Laub, M. T., Paulsen, I. T., Nelson, K. E., Eisen, J. A., Heidelberg, J. F., Alley, M. R., Ohta, N., Maddock, J. R., *et al*. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4136–4141.
9. Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. (1998) *Nucleic Acids Res.* **26,** 544–548.
10. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., *et al*. (1995) *Science* **269,** 496–512.
11. Waterman, M. S. (1988) *Methods Enzymol.* **164,** 765–793.
12. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000) *Nucleic Acids Res.* **28,** 263–266.
13. Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T. & White, O. (2001) *Nucleic Acids Res.* **29,** 41–43.
14. Riley, M. (1993) *Microbiol. Rev.* **57,** 862–952.
15. Karlin, S. (1998) *Curr. Opin. Microbiol.* **1,** 598–610.
16. Kim, H., Zhao, B., Snesrud, E. C., Haas, B. J., Town, C. D. & Quackenbush, J. (2002) *BioTechniques* **33,** 924–930.
17. Holden, M. T. G., Titball, R. W., Peacock, S. J., Cerdeño-Tárraga, A. M., Atkins, T., Crossman, L. C., Pitt, T., Churcher, C., Mungall, K., Bentley, S. D., *et al*. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 14240–14245.
18. DeShazer, D., Waag, D. M., Fritz, D. L. & Woods, D. E. (2001) *Microb. Pathogen.* **30,** 253–269.
19. Sanford, J. P. (1990) in *Principles and Practice of Infectious Diseases*, eds. Mandell, G. L., Douglas, R. G., Jr., & Bennett, J. E. (Churchill Livingstone, New York), pp. 1692–1696.
20. Bochner, B. R. (2003) *Nat. Rev. Genet.* **4,** 309–314.
21. Malakooti, J., Ely, B. & Matsumura, P. (1994) *J. Bacteriol.* **176,** 189–197.
22. DeShazer, D., Brett, P. J., Burtnick, M. N. & Woods, D. E. (1999) *J. Bacteriol.* **181,** 4661–4664.
23. Miller, W. R., Pannell, L., Cravitz, L., Tanner, W. A. & Ingalls, M. S. (1948) *J. Bacteriol.* **55,** 115–126.
24. Redfearn, M. S., Palleroni, N. J. & Stanier, R. Y. (1966) *J. Gen. Microbiol.* **43,** 293–313.
25. Moore, R. A., DeShazer, D., Reckseidler, S. L., Weissman, A. & Woods, D. E. (1999) *Antimicrob. Agents Chemother.* **43,** 465–470.
26. Bhatnagar, D., Ehrlich, K. C. & Cleveland, T. E. (2003) *Appl. Microbiol. Biotechnol.* **61,** 83–93.
27. Ulrich, R. L. & DeShazer, D. (2004) *Infect. Immun.* **72,** 1150–1154.
28. Powell, B. S., Powell, G. K., Morris, R. O., Rogowsky, P. M. & Kado, C. I. (1987) *Mol. Microbiol.* **1,** 309–316.
29. Miller, W. R., Pannell, L., Cravitz, L., Tanner, W. A. & Rosebury, T. (1948) *J. Bacteriol.* **55,** 127–135.